

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(43) 国际公布日:

2001年11月29日(29.11.01)

PCT

(10) 国际公布号:

WO 01/90369 A1

(51) 国际分类¹: C12N 15/54, 9/10, C07K 14/435, A61K 38/45

(21) 国际申请号: PCT/CN01/00749

(22) 国际申请日: 2001年5月14日(14.05.01)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
00115726.4 2000年5月16日(16.05.00) CN

(71) 申请人(对除美国以外的所有指定国): 上海博德基因开发有限公司(SHANGHAI BIOWINDOW GENE DEVELOPMENT INC.) [CN/CN]; 中国上海市中山北二路1111号3号楼12层, Shanghai 200092 (CN).

(72) 发明人: 及

(75) 发明人/申请人(仅对美国): 毛裕民(MAO, Yumin) [CN/CN]; 谢毅(XIE, Yi) [CN/CN]; 中国上海市中山北二路1111号3号楼12层, Shanghai 200092 (CN).

(74) 代理人: 上海华东专利事务所(SHANGHAI HUA DONG PATENT AGENCY); 中国上海市岳阳路319号, Shanghai 200031 (CN).

(81) 指定国(国家): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CO, CR, CU, CZ, DE,

DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW

(84) 指定国(地区): ARIPO专利(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), 欧亚专利(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲专利(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI专利(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG)

根据细则4.17的声明:

- 关于申请人在国际申请日有权申请并被授予专利(细则4.17(ii))对除美国以外的所有指定国
- 关于申请人在国际申请日有权要求该在先申请的优先权(细则4.17(iii))对除美国以外的所有指定国
- 关于申请人在国际申请日有权要求该在先申请的优先权(细则4.17(iii))对下列指定国: 美国
- 发明人资格(细则4.17(iv))仅对美国

本国际公布:

- 包括国际检索报告。

所引用双字母代码和其它缩写符号, 请参考刊登在每期PCT公报期刊起始的“代码及缩写符号简要说明”。

(54) Title: A NOVEL POLYPEPTIDE, A HUMAN N-ACETYL GALACTOSAMINE TRANSFERASE -28 AND THE POLYNUCLEOTIDE ENCODING THE POLYPEPTIDE

(54) 发明名称: 一种新的多肽—N-乙酰半乳糖胺转移酶-28 和编码这种多肽的多核苷酸

(57) Abstract: The present invention discloses a novel polypeptide, a human N-acetylgalactosamine transferase -28, the polynucleotide encoding the polypeptide and the method for producing the polypeptide by DNA recombinant technology. The invention also discloses the uses of the polypeptide in methods for treating various diseases, such as malignant tumour, hemopathy, developmental disorder, HIV infection, immunological disease, and various inflammations, etc. The invention also discloses the agonists against the polypeptide and the therapeutic action thereof. The invention also discloses the uses of the polynucleotide encoding the novel human N-acetylgalactosamine transferase -28.

(57) 摘要

本发明公开了一种新的多肽——N-乙酰半乳糖胺转移酶-28, 编码此多肽的多核苷酸和经 DNA 重组技术产生这种多肽的方法。本发明还公开了此多肽用于治疗多种疾病的方法, 如恶性肿瘤, 血液病, 发育紊乱症, HIV 感染, 免疫性疾病和各类炎症等。本发明还公开了抗此多肽的拮抗剂及其治疗作用。本发明还公开了编码这种新的 N-乙酰半乳糖胺转移酶-28 的多核苷酸的用途。

一种新的多肽——N-乙酰半乳糖胺转移酶-28 和编码这种多肽的多核苷酸

技术领域

本发明属于生物技术领域，具体地说，本发明描述了一种新的多肽——N-乙酰半乳糖胺转移酶-28，以及编码此多肽的多核苷酸序列。本发明还涉及此多核苷酸和多肽的制备方法和应用。

技术背景

O-糖苷键（粘蛋白-型）连接的寡糖分布于众多不同类型的糖蛋白中 [Sadler, J. E. John Wiley & Sons, 1984. *Biology of Carbohydrates* Vol. 2, pp. 199-213]。该类型的寡糖功能非常多样化，从特异性很强的寡糖功能，如参与细胞与细胞识别和寄主性病原体反应，到非常普通的寡糖功能，如保护蛋白质防止它的水解或为粘液分泌性蛋白提供合适的电荷与水结合环境 [Jentoft, N. 1990. *Trends Biochem. Sci.* 15, 291-294]。

O-糖苷键连接寡糖合成的起始反应是将一个 N-乙酰半乳糖胺基团从核苷糖 UDP-N-乙酰半乳糖胺转移到蛋白质受体的一个丝氨酸或一个苏氨酸残基上。该反应的催化酶是多肽 N-乙酰半乳糖胺转移酶（UDP-GalNAc），这个酶是位于分泌途径的细胞内酶。

最近生物学家从牛中克隆出了 N-乙酰半乳糖胺转移酶，该酶属于类型 II 膜蛋白，并同样包含了其他半乳糖转移酶的一般结构域。该酶的氨基酸序列包含 3 个一致序列作为 N 连接的糖基化位点。由于 N-乙酰半乳糖胺转移酶缺乏 N 末端 40 个氨基酸的膜结合片段和一个包括细胞质和跨膜结构域的序列片段，N-乙酰半乳糖胺转移酶在水中是可溶的 [Fred L. Homa, Tameira Hollander, Donna J. et al., 1993. *The Journal of Biological Chemistry* 268, 12609-12616]。

N-乙酰半乳糖胺转移酶的精确定位还存在着争议。有人认为 N-乙酰半乳糖在受体蛋白上的起始结合发生于前糙面内质网。而有的科学家则认为该结合是发生于后内质网组分或高尔基体顺面的后翻译事件。现在有证据表明，将 N-乙酰半乳糖胺转运至丝氨酸或苏氨酸的反应可能发生于分泌途径的几个相隔区域 [Schachter, H., and Brockhausen, I. 1992. Marcel Dekker, Inc., New York 263-332]。

本发明人的多肽与 N-乙酰半乳糖胺转移酶在蛋白质水平上有 34% 的相同性和 58% 的相似性，具有与 N-乙酰半乳糖胺转移酶相似的结构特征，属于 N-乙

酰半乳糖胺转移酶家族，故命名为 N-乙酰半乳糖胺转移酶-28，同时推测其具有相似的生物学功能。

由于如上所述 N-乙酰半乳糖胺转移酶-28 蛋白在调节细胞分裂和胚胎发育等机体重要功能中起重要作用，而且相信这些调节过程中涉及大量的蛋白，因而本领域中一直需要鉴定更多参与这些过程的 N-乙酰半乳糖胺转移酶-28 蛋白，特别是鉴定这种蛋白的氨基酸序列。新 N-乙酰半乳糖胺转移酶-28 蛋白编码基因的分离也为研究确定该蛋白在健康和疾病状态下的作用提供了基础。这种蛋白可能构成开发疾病诊断和/或治疗药的基础，因此分离其编码 DNA 是非常重要的。

发明目的

本发明的一个目的是提供分离的新的多肽——N-乙酰半乳糖胺转移酶-28 以及其片段、类似物和衍生物。

本发明的另一个目的是提供编码该多肽的多核苷酸。

本发明的另一个目的是提供含有编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸的重组载体。

本发明的另一个目的是提供含有编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸的基因工程化宿主细胞。

本发明的另一个目的是提供生产 N-乙酰半乳糖胺转移酶-28 的方法。

本发明的另一个目的是提供针对本发明的多肽——N-乙酰半乳糖胺转移酶-28 的抗体。

本发明的另一个目的是提供了针对本发明多肽——N-乙酰半乳糖胺转移酶-28 的模拟化合物、拮抗剂、激动剂、抑制剂。

本发明的另一个目的是提供诊断治疗与 N-乙酰半乳糖胺转移酶-28 异常相关的疾病的方法。

发明概要

本发明涉及一种分离的多肽，该多肽是人源的，它包含：具有 SEQ ID No. 2 氨基酸序列的多肽、或其保守性变体、生物活性片段或衍生物。较佳地，该多肽是具有 SEQ ID NO: 2 氨基酸序列的多肽。

本发明还涉及一种分离的多核苷酸，它包含选自下组的一种核苷酸序列或其变体：

- (a) 编码具有 SEQ ID No. 2 氨基酸序列的多肽的多核苷酸;
- (b) 与多核苷酸 (a) 互补的多核苷酸;
- (c) 与 (a) 或 (b) 的多核苷酸序列具有至少 70% 相同性的多核苷酸。

更佳地, 该多核苷酸的序列是选自下组的一种: (a) 具有 SEQ ID NO: 1 中 193-966 位的序列; 和 (b) 具有 SEQ ID NO: 1 中 1-1271 位的序列。

本发明另外涉及一种含有本发明多核苷酸的载体, 特别是表达载体; 一种用该载体遗传工程化的宿主细胞, 包括转化、转导或转染的宿主细胞; 一种包括培养所述宿主细胞和回收表达产物的制备本发明多肽的方法。

本发明还涉及一种能与本发明多肽特异性结合的抗体。

本发明还涉及一种筛选的模拟、激活、拮抗或抑制 N-乙酰半乳糖胺转移酶-28 蛋白活性的化合物的方法, 其包括利用本发明的多肽。本发明还涉及用该方法获得的化合物。

本发明还涉及一种体外检测与 N-乙酰半乳糖胺转移酶-28 蛋白异常表达相关的疾病或疾病易感性的方法, 包括检测生物样品中所述多肽或其编码多核苷酸序列中的突变, 或者检测生物样品中本发明多肽的量或生物活性。

本发明也涉及一种药物组合物, 它含有本发明多肽或其模拟物、激活剂、拮抗剂或抑制剂以及药学上可接受的载体。

本发明还涉及本发明的多肽和/或多核苷酸在制备用于治疗癌症、发育性疾病或免疫性疾病或其它由于 N-乙酰半乳糖胺转移酶-28 表达异常所引起疾病的药物的用途。

本发明的其它方面由于本文的技术的公开, 对本领域的技术人员而言是显而易见的。

附图说明

下列附图用于说明本发明的具体实施方案, 而不用于限定由权利要求书所界定的本发明范围。

图1是本发明 N-乙酰半乳糖胺转移酶-28 和 N-乙酰半乳糖胺转移酶的氨基酸序列同源性比较图。上方序列是 N-乙酰半乳糖胺转移酶-28, 下方序列是 N-乙酰半乳糖胺转移酶。相同氨基酸在两个序列间用单字符氨基酸表示, 相似氨基酸用 “+” 表示。

图 2 为分离的 N-乙酰半乳糖胺转移酶-28 的聚丙烯酰胺凝胶电泳图 (SDS-PAGE)。28kDa 为蛋白质的分子量。箭头所指为分离出的蛋白条带。

发明内容

本说明书和权利要求书中使用的下列术语除非特别说明具有如下的含义：

“核酸序列”是指寡核苷酸、核苷酸或多核苷酸及其片段或部分，也可以指基因组或合成的DNA或RNA，它们可以是单链或双链的，代表有义链或反义链。类似地，术语“氨基酸序列”是指寡肽、肽、多肽或蛋白质序列及其片段或部分。当本发明中的“氨基酸序列”涉及一种天然存在的蛋白质分子的氨基酸序列时，这种“多肽”或“蛋白质”不意味着将氨基酸序列限制为与所述蛋白质分子相关的完整的天然氨基酸。

蛋白质或多核苷酸“变体”是指一种具有一个或多个氨基酸或核苷酸改变的氨基酸序列或编码它的多核苷酸序列。所述改变可包括氨基酸序列或核苷酸序列中氨基酸或核苷酸的缺失、插入或替换。变体可具有“保守性”改变，其中替换的氨基酸具有与原氨基酸相类似的结构或化学性质，如用亮氨酸替换异亮氨酸。变体也可具有非保守性改变，如用色氨酸替换甘氨酸。

“缺失”是指在氨基酸序列或核苷酸序列中一个或多个氨基酸或核苷酸的缺失。

“插入”或“添加”是指在氨基酸序列或核苷酸序列中的改变导致与天然存在的分子相比，一个或多个氨基酸或核苷酸的增加。“替换”是指由不同的氨基酸或核苷酸替换一个或多个氨基酸或核苷酸。

“生物活性”是指具有天然分子的结构、调控或生物化学功能的蛋白质。类似地，术语“免疫学活性”是指天然的、重组的或合成蛋白质及其片段在合适的动物或细胞中诱导特定免疫反应以及与特异性抗体结合的能力。

“激动剂”是指当与N-乙酰半乳糖胺转移酶-28结合时，一种可引起该蛋白质改变从而调节该蛋白质活性的分子。激动剂可以包括蛋白质、核酸、碳水化合物或任何其它可结合N-乙酰半乳糖胺转移酶-28的分子。

“拮抗剂”或“抑制物”是指当与N-乙酰半乳糖胺转移酶-28结合时，一种可封闭或调节N-乙酰半乳糖胺转移酶-28的生物学活性或免疫学活性的分子。拮抗剂和抑制物可以包括蛋白质、核酸、碳水化合物或任何其它可结合N-乙酰半乳糖胺转移酶-28的分子。

“调节”是指N-乙酰半乳糖胺转移酶-28的功能发生改变，包括蛋白质活性的升高或降低、结合特性的改变及N-乙酰半乳糖胺转移酶-28的任何其它生物学性质、功能或免疫性质的改变。

“基本上纯”是指基本上不含天然与其相关的其它蛋白、脂类、糖类或其它物

质。本领域的技术人员能用标准的蛋白质纯化技术纯化 N-乙酰半乳糖胺转移酶-28。基本上纯的 N-乙酰半乳糖胺转移酶-28 在非还原性聚丙烯酰胺凝胶上能产生单一的主带。N-乙酰半乳糖胺转移酶-28 多肽的纯度可用氨基酸序列分析。

“互补的”或“互补”是指在允许的盐浓度和温度条件下通过碱基配对的多核苷酸天然结合。例如，序列“C-T-G-A”可与互补的序列“G-A-C-T”结合。两个单链分子之间的互补可以是部分的或全部的。核酸链之间的互补程度对于核酸链之间杂交的效率及强度有明显影响。

“同源性”是指互补的程度，可以是部分同源或完全同源。“部分同源”是指一种部分互补的序列，其至少可部分抑制完全互补的序列与靶核酸的杂交。这种杂交的抑制可通过在严格性程度降低的条件下进行杂交（Southern印迹或Northern印迹等）来检测。基本上同源的序列或杂交探针可竞争和抑制完全同源的序列与靶序列在严格性程度降低的条件下的结合。这并不意味着严格性程度降低的条件允许非特异性结合，因为严格性程度降低的条件要求两条序列相互的结合为特异性或选择性相互作用。

“相同性百分率”是指在两种或多种氨基酸或核酸序列比较中序列相同或相似的百分率。可用电子方法测定相同性百分率，如通过MEGALIGN程序（Lasergene software package, DNASTAR, Inc., Madison Wis.）。MEGALIGN程序可根据不同的方法如Cluster法比较两种或多种序列（Higgins, D. G. 和 P.M. Sharp (1988) Gene 73:237-244）。Cluster法通过检查所有配对之间的距离将各组序列排列成簇。然后将各簇以成对或成组分配。两个氨基酸序列如序列A和序列B之间的相同性百分率通过下式计算：

$$\frac{\text{序列A与序列B之间匹配的残基数}}{\text{序列A的残基数} - \text{序列A中间隔残基数} - \text{序列B中间隔残基数}} \times 100$$

也可以通过Cluster法或用本领域周知的方法如Jotun Hein 测定核酸序列之间的相同性百分率（Hein J., (1990) Methods in enzymology 183:625-645）。

“相似性”是指氨基酸序列之间排列对比时相应位置氨基酸残基的相同或保守性取代的程度。用于保守性取代的氨基酸，例如带负电荷的氨基酸可包括天冬氨酸和谷氨酸；带正电荷的氨基酸可包括赖氨酸和精氨酸；具有不带电荷的头部基团有相似亲水性的氨基酸可包括亮氨酸、异亮氨酸和缬氨酸；甘氨酸和丙氨酸；天冬酰胺和谷氨酰胺；丝氨酸和苏氨酸；苯丙氨酸和酪氨酸。

“反义”是指与特定的DNA或RNA序列互补的核苷酸序列。“反义链”是指与“有义链”互补的核酸链。

“衍生物”是指HFP或编码其核酸的化学修饰物。这种化学修饰物可以用烷基、酰基或氨基替换氢原子。核酸衍生物可编码保留天然分子的主要生物学特性的多肽。

“抗体”是指完整的抗体分子及其片段，如 Fa 、 $F(ab')_2$ 及 Fv ，其能特异性结合N-乙酰半乳糖胺转移酶-28的抗原决定簇。

“人源化抗体”是指非抗原结合区域的氨基酸序列被替换变得与人抗体更为相似，但仍保留原始结合活性的抗体。

“分离的”一词指将物质从它原来的环境（例如，若是自然产生的就指其天然环境）之中移出。比如说，一个自然产生的多核苷酸或多肽存在于活动物中就是没有被分离出来，但同样的多核苷酸或多肽同一些或全部在自然系统中与之共存的物质分开就是分离的。这样的多核苷酸可能是某一载体的一部分，也可能这样的多核苷酸或多肽是某一组合物的一部分。既然载体或组合物不是它天然环境的成分，它们仍然是分离的。

如本发明所用，“分离的”是指物质从其原始环境中分离出来（如果是天然物质，原始环境即是天然环境）。如活体细胞内的天然状态下的多聚核苷酸和多肽是没有分离纯化的，但同样的多聚核苷酸或多肽如从天然状态中同存在的其他物质中分开，则为分离纯化的。

如本文所用，“分离的 N-乙酰半乳糖胺转移酶-28”是指 N-乙酰半乳糖胺转移酶-28 基本上不含天然与其相关的其它蛋白、脂类、糖类或其它物质。本领域的技术人员能用标准的蛋白质纯化技术纯化 N-乙酰半乳糖胺转移酶-28。基本上纯的多肽在非还原聚丙烯酰胺凝胶上能产生单一的主带。N-乙酰半乳糖胺转移酶-28 多肽的纯度能用氨基酸序列分析。

本发明提供了一种新的多肽——N-乙酰半乳糖胺转移酶-28，其基本上是由SEQ ID NO: 2所示的氨基酸序列组成的。本发明的多肽可以是重组多肽、天然多肽、合成多肽，优选重组多肽。本发明的多肽可以是天然纯化的产物，或是化学合成的产物，或使用重组技术从原核或真核宿主（例如，细菌、酵母、高等植物、昆虫和哺乳动物细胞）中产生。根据重组生产方案所用的宿主，本发明的多肽可以是糖基化的，或可以是非糖基化的。本发明的多肽还可包括或不包括起始的甲硫氨酸残基。

本发明还包括 N-乙酰半乳糖胺转移酶-28 的片段、衍生物和类似物。如本发明所用，术语“片段”、“衍生物”和“类似物”是指基本上保持本发明的 N-乙酰半乳糖胺转移酶-28 相同的生物学功能或活性的多肽。本发明多肽的片

段、衍生物或类似物可以是：(I) 这样一种，其中一个或多个氨基酸残基被保守或非保守氨基酸残基（优选的是保守氨基酸残基）取代，并且取代的氨基酸可以是也可以不是由遗传密码子编码的；或者(II) 这样一种，其中一个或多个氨基酸残基上的某个基团被其它基团取代包含取代基；或者(III) 这样一种，其中成熟多肽与另一种化合物（比如延长多肽半衰期的化合物，例如聚乙二醇）融合；或者(IV) 这样一种，其中附加的氨基酸序列融合进成熟多肽而形成的多肽序列（如前导序列或分泌序列或用来纯化此多肽的序列或蛋白原序列）。通过本文的阐述，这样的片段、衍生物和类似物被认为在本领域技术人员知识范围之内。

本发明提供了分离的核酸（多核苷酸），基本由编码具有 SEQ ID NO: 2 氨基酸序列的多肽的多核苷酸组成。本发明的多核苷酸序列包括 SEQ ID NO: 1 的核苷酸序列。本发明的多核苷酸是从人胎脑组织的 cDNA 文库中发现的。它包含的多核苷酸序列全长为 1271 个碱基，其开放读框 193-966 编码了 257 个氨基酸。根据氨基酸序列同源比较发现，此多肽与 N-乙酰半乳糖胺转移酶有 34% 的同源性，可推断出该 N-乙酰半乳糖胺转移酶-28 具有 N-乙酰半乳糖胺转移酶相似的结构和功能。

本发明的多核苷酸可以是 DNA 形式或是 RNA 形式。DNA 形式包括 cDNA、基因组 DNA 或人工合成的 DNA。DNA 可以是单链的或是双链的。DNA 可以是编码链或非编码链。编码成熟多肽的编码区序列可以与 SEQ ID NO: 1 所示的编码区序列相同或者是简并的变异体。如本发明所用，“简并的变异体”在本发明中是指编码具有 SEQ ID NO: 2 的蛋白质或多肽，但与 SEQ ID NO: 1 所示的编码区序列有差别的核酸序列。

编码 SEQ ID NO: 2 的成熟多肽的多核苷酸包括：只有成熟多肽的编码序列；成熟多肽的编码序列和各种附加编码序列；成熟多肽的编码序列（和任选的附加编码序列）以及非编码序列。

术语“编码多肽的多核苷酸”是指包括编码此多肽的多核苷酸和包括附加编码和/或非编码序列的多核苷酸。

本发明还涉及上述描述多核苷酸的变异体，其编码与本发明有相同的氨基酸序列的多肽或多肽的片断、类似物和衍生物。此多核苷酸的变异体可以是天然发生的等位变异体或非天然发生的变异体。这些核苷酸变异体包括取代变异体、缺失变异体和插入变异体。如本领域所知的，等位变异体是一个多核苷酸的替换形式，它可能是一个或多个核苷酸的取代、缺失或插入，但不会从实质

上改变其编码的多肽的功能。

本发明还涉及与以上所描述的序列杂交的多核苷酸（两个序列之间具有至少 50%，优选具有 70% 的相同性）。本发明特别涉及在严格条件下与本发明所述多核苷酸可杂交的多核苷酸。在本发明中，“严格条件”是指：(1) 在较低离子强度和较高温度下的杂交和洗脱，如 0.2×SSC，0.1%SDS，60℃；或(2) 杂交时加用变性剂，如 50%(v/v) 甲酰胺，0.1%小牛血清/0.1%Ficoll，42℃等；或(3) 仅在两条序列之间的相同性至少在 95%以上，更好是 97%以上时才发生杂交。并且，可杂交的多核苷酸编码的多肽与 SEQ ID NO: 2 所示的成熟多肽有相同的生物学功能和活性。

本发明还涉及与以上所描述的序列杂交的核酸片段。如本发明所用，“核酸片段”的长度至少含 10 个核苷酸，较好是至少 20-30 个核苷酸，更好是至少 50-60 个核苷酸，最好是至少 100 个核苷酸以上。核酸片段也可用于核酸的扩增技术(如 PCR) 以确定和/或分离编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸。

本发明中的多肽和多核苷酸优选以分离的形式提供，更佳地被纯化至均质。

本发明的编码 N-乙酰半乳糖胺转移酶-28 的特异的多核苷酸序列能用多种方法获得。例如，用本领域熟知的杂交技术分离多核苷酸。这些技术包括但不限于：1) 用探针与基因组或 cDNA 文库杂交以检出同源的多核苷酸序列，和 2) 表达文库的抗体筛选以检出具有共同结构特征的克隆的多核苷酸片段。

本发明的 DNA 片段序列也能用下列方法获得：1) 从基因组 DNA 分离双链 DNA 序列；2) 化学合成 DNA 序列以获得所述多肽的双链 DNA。

上述提到的方法中，分离基因组 DNA 最不常用。DNA 序列的直接化学合成是经常选用的方法。更经常选用的方法是 cDNA 序列的分离。分离感兴趣的 cDNA 的标准方法是从高表达该基因的供体细胞分离 mRNA 并进行逆转录，形成质粒或噬菌体 cDNA 文库。提取 mRNA 的方法已有多种成熟的技术，试剂盒也可从商业途径获得(Qiagene)。而构建 cDNA 文库也是通常的方法(Sambrook, et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, New York, 1989)。还可得到商业供应的 cDNA 文库，如 Clontech 公司的不同 cDNA 文库。当结合使用聚合酶反应技术时，即使极少的表达产物也能克隆。

可用常规方法从这些 cDNA 文库中筛选本发明的基因。这些方法包括(但不限于)：(1) DNA-DNA 或 DNA-RNA 杂交；(2) 标志基因功能的出现或丧失；(3) 测定 N-乙酰半乳糖胺转移酶-28 的转录本的水平；(4) 通过免疫学技术或测定生物学活性，来检测基因表达的蛋白产物。上述方法可单用，也可多种方法联合

应用。

在第(1)种方法中，杂交所用的探针是与本发明的多核苷酸的任何一部分同源，其长度至少 10 个核苷酸，较好是至少 30 个核苷酸，更好是至少 50 个核苷酸，最好是至少 100 个核苷酸。此外，探针的长度通常在 2000 个核苷酸之内，较佳的为 1000 个核苷酸之内。此处所用的探针通常是在本发明的基因序列信息的基础上化学合成的 DNA 序列。本发明的基因本身或者片段当然可以用作探针。DNA 探针的标记可用放射性同位素，荧光素或酶(如碱性磷酸酶)等。

在第(4)种方法中，检测 N-乙酰半乳糖胺转移酶-28 基因表达的蛋白产物可用免疫学技术如 Western 印迹法、放射免疫沉淀法、酶联免疫吸附法(ELISA)等。

应用 PCR 技术扩增 DNA/RNA 的方法(Saiki, et al. Science 1985; 230:1350-1354)被优选用于获得本发明的基因。特别是很难从文库中得到全长的 cDNA 时，可优选使用 RACE 法(RACE - cDNA 末端快速扩增法)，用于 PCR 的引物可根据本文所公开的本发明的多核苷酸序列信息适当地选择，并可用常规方法合成。可用常规方法如通过凝胶电泳分离和纯化扩增的 DNA/RNA 片段。

如上所述得到的本发明的基因，或者各种 DNA 片段等的多核苷酸序列可用常规方法如双脱氧链终止法(Sanger et al. PNAS, 1977, 74: 5463-5467)测定。这类多核苷酸序列测定也可用商业测序试剂盒等。为了获得全长的 cDNA 序列，测序需反复进行。有时需要测定多个克隆的 cDNA 序列，才能拼接成全长的 cDNA 序列。

本发明也涉及包含本发明的多核苷酸的载体，以及用本发明的载体或直接用 N-乙酰半乳糖胺转移酶-28 编码序列经基因工程产生的宿主细胞，以及经重组技术产生本发明所述多肽的方法。

本发明中，编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸序列可插入到载体中，以构成含有本发明所述多核苷酸的重组载体。术语“载体”指本领域熟知的细菌质粒、噬菌体、酵母质粒、植物细胞病毒、哺乳动物细胞病毒如腺病毒、逆转录病毒或其它载体。在本发明中适用的载体包括但不限于：在细菌中表达的基于 T7 启动子的表达载体(Rosenberg, et al. Gene, 1987, 56:125)；在哺乳动物细胞中表达的 pMSXND 表达载体(Lee and Nathans, J Bio Chem. 263: 3521, 1988)和在昆虫细胞中表达的来源于杆状病毒的载体。总之，只要能在宿主体内复制和稳定，任何质粒和载体都可以用于构建重组表达载体。表达载体的一个重要特征是通常含有复制起始点、启动子、标记基因和翻译调控元

件。

本领域的技术人员熟知的方法能用于构建含编码 N-乙酰半乳糖胺转移酶-28 的 DNA 序列和合适的转录/翻译调控元件的表达载体。这些方法包括体外重组 DNA 技术、DNA 合成技术、体内重组技术等 (Sambrook, et al. *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory, New York, 1989)。所述的 DNA 序列可有效连接到表达载体中的适当启动子上, 以指导 mRNA 合成。这些启动子的代表性例子有: 大肠杆菌的 lac 或 trp 启动子; λ 噬菌体的 PL 启动子; 真核启动子包括 CMV 立即早期启动子、HSV 胸苷激酶启动子、早期和晚期 SV40 启动子、反转录病毒的 LTRs 和其它一些已知的可控制基因在原核细胞或真核细胞或其病毒中表达的启动子。表达载体还包括翻译起始用的核糖体结合位点和转录终止子等。在载体中插入增强子序列将会使其在高等真核细胞中的转录得到增强。增强子是 DNA 表达的顺式作用因子, 通常大约有 10 到 300 个碱基对, 作用于启动子以增强基因的转录。可举的例子包括在复制起始点晚期一侧的 100 到 270 个碱基对的 SV40 增强子、在复制起始点晚期一侧的多瘤增强子以及腺病毒增强子等。

此外, 表达载体优选地包含一个或多个选择性标记基因, 以提供用于选择转化的宿主细胞的表型性状, 如真核细胞培养用的二氢叶酸还原酶、新霉素抗性以及绿色荧光蛋白 (GFP), 或用于大肠杆菌的四环素或氨苄青霉素抗性等。

本领域一般技术人员都清楚如何选择适当的载体/转录调控元件 (如启动子、增强子等) 和选择性标记基因。

本发明中, 编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸或含有该多核苷酸的重组载体可转化或转导入宿主细胞, 以构成含有该多核苷酸或重组载体的基因工程化宿主细胞。术语“宿主细胞”指原核细胞, 如细菌细胞; 或是低等真核细胞, 如酵母细胞; 或是高等真核细胞, 如哺乳动物细胞。代表性例子有: 大肠杆菌, 链霉菌属; 细菌细胞如鼠伤寒沙门氏菌; 真菌细胞如酵母; 植物细胞; 昆虫细胞如果蝇 S2 或 Sf9; 动物细胞如 CHO、COS 或 Bowes 黑素瘤细胞等。

用本发明所述的 DNA 序列或含有所述 DNA 序列的重组载体转化宿主细胞可用本领域技术人员熟知的常规技术进行。当宿主为原核生物如大肠杆菌时, 能吸收 DNA 的感受态细胞可在指数生长期后收获, 用 CaCl_2 法处理, 所用的步骤在本领域众所周知。可供选择的是用 MgCl_2 。如果需要, 转化也可用电穿孔的方法进行。当宿主是真核生物, 可选用如下的 DNA 转染方法: 磷酸钙共沉淀法, 或者常规机械方法如显微注射、电穿孔、脂质体包装等。

通过常规的重组 DNA 技术, 利用本发明的多核苷酸序列可用来表达或生产重组的 N-乙酰半乳糖胺转移酶-28 (Science, 1984; 224: 1431)。一般来说有以下步骤:

- (1) 用本发明的编码人 N-乙酰半乳糖胺转移酶-28 的多核苷酸(或变异性), 或用含有该多核苷酸的重组表达载体转化或转导合适的宿主细胞;
- (2) 在合适的培养基中培养宿主细胞;
- (3) 从培养基或细胞中分离、纯化蛋白质。

在步骤(2)中, 根据所用的宿主细胞, 培养中所用的培养基可选自各种常规培养基。在适于宿主细胞生长的条件下进行培养。当宿主细胞生长到适当的细胞密度后, 用合适的方法(如温度转换或化学诱导)诱导选择的启动子, 将细胞再培养一段时间。

在步骤(3)中, 重组多肽可包被于细胞内、或在细胞膜上表达、或分泌到细胞外。如果需要, 可利用其物理的、化学的和其它特性通过各种分离方法分离和纯化重组的蛋白。这些方法是本领域技术人员所熟知的。这些方法包括但不限于: 常规的复性处理、蛋白沉淀剂处理(盐析方法)、离心、渗透破菌、超声波处理、超离心、分子筛层析(凝胶过滤)、吸附层析、离子交换层析、高效液相层析(HPLC)和其它各种液相层析技术及这些方法的结合。

本发明的多肽以及该多肽的拮抗剂、激动剂和抑制剂可直接用于疾病治疗, 例如, 可治疗恶性肿瘤、肾上腺缺乏症、皮肤病、各类炎症、HIV 感染和免疫性疾病等。

O-糖苷键(粘蛋白-型)连接的寡糖分布于众多不同类型的糖蛋白中。该类型的寡糖功能非常多样化, 如参与细胞与细胞识别和寄主性病原体反应的特异性很强的寡糖, 如保护蛋白质防止它的水解或为粘液分泌性蛋白提供合适的电荷与水结合环境的寡糖。

O-糖苷键连接寡糖合成的起始反应需要催化酶多肽 N-乙酰半乳糖胺转移酶(UDP-GalNAc), 它是位于分泌途径的细胞内酶。

本发明的多肽与 N-乙酰半乳糖胺转移酶是 N-乙酰半乳糖胺转移酶, 含 N-乙酰半乳糖胺转移酶家族的特征性序列, 两者具有相似的生物学功能。它在体内参与 O-糖苷键连接寡糖合成的起始反应, 是合成 O-糖苷键连接寡糖所必需, 而 O-糖苷键连接寡糖对于多种糖蛋白功能起着重要作用, 其表达异常通常与多种糖蛋白功能异常密切相关, 影响免疫调节、炎症反应、细胞分裂等, 并产生相关的疾病。

由此可见，本发明的 N-乙酰半乳糖胺转移酶-28 的表达异常将产生各种疾病尤其是各种肿瘤、胚胎发育紊乱症、生长发育障碍性疾病、炎症、免疫性疾病，这些疾病包括但不限于：

各种组织的肿瘤：胃癌，肝癌，肺癌，食管癌，乳腺癌，白血病，淋巴瘤，甲状腺肿瘤，子宫肌瘤，神经细胞瘤，星形细胞瘤，室管膜瘤，胶质细胞瘤，神经纤维瘤，结肠癌，黑色素瘤，膀胱癌，子宫癌，子宫内膜癌，胸腺肿瘤，鼻咽癌，喉癌，气管肿瘤，纤维瘤，纤维肉瘤，脂肪瘤，脂肪肉瘤

胚胎发育紊乱症：先天性流产，腭裂，肢体缺如，肢体分化障碍，房间隔缺损，神经管缺陷，先天性脑积水，先天性青光眼或白内障，先天性耳聋

生长发育障碍性疾病：精神发育迟缓，脑发育障碍，皮肤、脂肪和肌肉发育不良性疾病，骨与关节发育不良性疾病，各种代谢缺陷病，呆小症，侏儒症，库兴综合症，性发育迟缓症

炎症：慢性活动性肝炎，结节病，多肌炎，慢性鼻炎，慢性胃炎，脑脊髓多发性硬化，肾小球性肾炎，心肌炎，心肌病，动脉粥样硬化，胃溃疡，宫颈炎，各种感染性炎症

免疫性疾病：系统性红斑狼疮，类风湿性关节炎，支气管哮喘，荨麻疹，特异性皮炎，感染后心肌炎，硬皮病，重症肌无力，格林-巴利综合症，普通易变免疫缺陷病，原发性 B 淋巴细胞免疫缺陷病，获得性免疫缺陷综合症

本发明的 N-乙酰半乳糖胺转移酶-28 的表达异常还将产生某些遗传性，血液性疾病等。

本发明的多肽以及该多肽的拮抗剂、激动剂和抑制剂可直接用于疾病治疗，例如，可治疗各种疾病尤其是各种肿瘤、胚胎发育紊乱症、生长发育障碍性疾病、炎症、免疫性疾病，某些遗传性，血液性疾病等。

本发明也提供了筛选化合物以鉴定提高(激动剂)或阻遏(拮抗剂)N-乙酰半乳糖胺转移酶-28 的药剂的方法。激动剂提高 N-乙酰半乳糖胺转移酶-28 刺激细胞增殖等生物功能，而拮抗剂阻止和治疗与细胞过度增殖有关的紊乱如各种癌症。例如，能在药物的存在下，将哺乳动物细胞或表达 N-乙酰半乳糖胺转移酶-28 的膜制剂与标记的 N-乙酰半乳糖胺转移酶-28 一起培养。然后测定药物提高或阻遏此相互作用的能力。

N-乙酰半乳糖胺转移酶-28 的拮抗剂包括筛选出的抗体、化合物、受体缺失物和类似物等。N-乙酰半乳糖胺转移酶-28 的拮抗剂可以与 N-乙酰半乳糖胺转移酶-28 结合并消除其功能，或是抑制该多肽的产生，或是与该多肽的活性

位点结合使该多肽不能发挥生物学功能。

在筛选作为拮抗剂的化合物时，可以将 N-乙酰半乳糖胺转移酶-28 加入生物分析测定中，通过测定化合物对 N-乙酰半乳糖胺转移酶-28 和其受体之间相互作用的影响来确定化合物是否是拮抗剂。用上述筛选化合物的同样方法，可以筛选出起拮抗剂作用的受体缺失物和类似物。能与 N-乙酰半乳糖胺转移酶-28 结合的多肽分子可通过筛选由各种可能组合的氨基酸结合于固相物组成的随机多肽库而获得。筛选时，一般应对 N-乙酰半乳糖胺转移酶-28 分子进行标记。

本发明提供了用多肽，及其片段、衍生物、类似物或它们的细胞作为抗原以生产抗体的方法。这些抗体可以是多克隆抗体或单克隆抗体。本发明还提供了针对 N-乙酰半乳糖胺转移酶-28 抗原决定簇的抗体。这些抗体包括(但不限于)：多克隆抗体、单克隆抗体、嵌合抗体、单链抗体、Fab 片段和 Fab 表达文库产生的片段。

多克隆抗体的生产可用 N-乙酰半乳糖胺转移酶-28 直接注射免疫动物(如家兔，小鼠，大鼠等)的方法得到，多种佐剂可用于增强免疫反应，包括但不限于弗氏佐剂等。制备 N-乙酰半乳糖胺转移酶-28 的单克隆抗体的技术包括但不限于杂交瘤技术(Kohler and Milstein, *Nature*, 1975, 256: 495-497)，三瘤技术，人 B-细胞杂交瘤技术，EBV-杂交瘤技术等。将人恒定区和非人源的可变区结合的嵌合抗体可用已有的技术生产(Morrison et al, *PNAS*, 1985, 81: 6851)。而已有的生产单链抗体的技术(U.S. Pat No. 4946778)也可用于生产抗 N-乙酰半乳糖胺转移酶-28 的单链抗体。

抗 N-乙酰半乳糖胺转移酶-28 的抗体可用于免疫组织化学技术中，检测活检标本中的 N-乙酰半乳糖胺转移酶-28。

与 N-乙酰半乳糖胺转移酶-28 结合的单克隆抗体也可用放射性同位素标记，注入体内可跟踪其位置和分布。这种放射性标记的抗体可作为一种非创伤性诊断方法用于肿瘤细胞的定位和判断是否有转移。

抗体还可用于设计针对体内某一特殊部位的免疫毒素。如 N-乙酰半乳糖胺转移酶-28 高亲和性的单克隆抗体可与细菌或植物毒素(如白喉毒素，蓖麻蛋白，红豆碱等)共价结合。一种通常的方法是用巯基交联剂如 SPDP，攻击抗体的氨基，通过二硫键的交换，将毒素结合于抗体上，这种杂交抗体可用于杀灭 N-乙酰半乳糖胺转移酶-28 阳性的细胞。

本发明中的抗体可用于治疗或预防与 N-乙酰半乳糖胺转移酶-28 相关的疾病。给予适当剂量的抗体可以刺激或阻断 N-乙酰半乳糖胺转移酶-28 的产生或

活性。

本发明还涉及定量和定位检测 N-乙酰半乳糖胺转移酶-28 水平的诊断试验方法。这些试验是本领域所熟知的，且包括 FISH 测定和放射免疫测定。试验中所检测的 N-乙酰半乳糖胺转移酶-28 水平，可以用作解释 N-乙酰半乳糖胺转移酶-28 在各种疾病中的重要性和用于诊断 N-乙酰半乳糖胺转移酶-28 起作用的疾病。

本发明的多肽还可用作肽谱分析，例如，多肽可用物理的、化学或酶进行特异性切割，并进行一维或二维或三维的凝胶电泳分析，更好的是进行质谱分析。

编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸也可用于多种治疗目的。基因治疗技术可用于治疗由于 N-乙酰半乳糖胺转移酶-28 的无表达或异常/无活性表达所致的细胞增殖、发育或代谢异常。重组的基因治疗载体(如病毒载体)可设计用于表达变异的 N-乙酰半乳糖胺转移酶-28，以抑制内源性的 N-乙酰半乳糖胺转移酶-28 活性。例如，一种变异的 N-乙酰半乳糖胺转移酶-28 可以是缩短的、缺失了信号传导功能域的 N-乙酰半乳糖胺转移酶-28，虽可与下游的底物结合，但缺乏信号传导活性。因此，重组的基因治疗载体可用于治疗 N-乙酰半乳糖胺转移酶-28 表达或活性异常所致的疾病。来源于病毒的表达载体如逆转录病毒、腺病毒、腺病毒相关病毒、单纯疱疹病毒、细小病毒等可用于将编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸转移至细胞内。构建携带编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸的重组病毒载体的方法可见于已有文献(Sambrook, et al.)。另外，重组编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸可包装到脂质体中转移至细胞内。

多核苷酸导入组织或细胞内的方法包括：将多核苷酸直接注入到体内组织中；或在体外通过载体(如病毒、噬菌体或质粒等)先将多核苷酸导入细胞中，再将细胞移植到体内等。

抑制 N-乙酰半乳糖胺转移酶-28 mRNA 的寡核苷酸(包括反义 RNA 和 DNA)以及核酶也在本发明的范围之内。核酶是一种能特异性分解特定 RNA 的酶样 RNA 分子，其作用机制是核酶分子与互补的靶 RNA 特异性杂交后进行核酸内切作用。反义的 RNA 和 DNA 及核酶可用已有的任何 RNA 或 DNA 合成技术获得，如固相磷酸酰胺化学合成法合成寡核苷酸的技术已广泛应用。反义 RNA 分子可通过编码该 RNA 的 DNA 序列在体外或体内转录获得。这种 DNA 序列已整合到载体的 RNA 聚合酶启动子的下游。为了增加核酸分子的稳定性，可用多种方法对其进行修

饰，如增加两侧的序列长度，核糖核苷之间的连接应用磷酸硫酯键或肽键而非磷酸二酯键。

编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸可用于与 N-乙酰半乳糖胺转移酶-28 的相关疾病的诊断。编码 N-乙酰半乳糖胺转移酶-28 的多核苷酸可用于检测 N-乙酰半乳糖胺转移酶-28 的表达与否或在疾病状态下 N-乙酰半乳糖胺转移酶-28 的异常表达。如编码 N-乙酰半乳糖胺转移酶-28 的 DNA 序列可用于对活检标本进行杂交以判断 N-乙酰半乳糖胺转移酶-28 的表达状况。杂交技术包括 Southern 印迹法、Northern 印迹法、原位杂交等。这些技术方法都是公开的成熟技术，相关的试剂盒都可从商业途径得到。本发明的多核苷酸的一部分或全部可作为探针固定在微阵列 (Microarray) 或 DNA 芯片 (又称为“基因芯片”) 上，用于分析组织中基因的差异表达分析和基因诊断。用 N-乙酰半乳糖胺转移酶-28 特异的引物进行 RNA-聚合酶链反应 (RT-PCR) 体外扩增也可检测 N-乙酰半乳糖胺转移酶-28 的转录产物。

检测 N-乙酰半乳糖胺转移酶-28 基因的突变也可用于诊断 N-乙酰半乳糖胺转移酶-28 相关的疾病。N-乙酰半乳糖胺转移酶-28 突变的形式包括与正常野生型 N-乙酰半乳糖胺转移酶-28 DNA 序列相比的点突变、易位、缺失、重组和其它任何异常等。可用已有的技术如 Southern 印迹法、DNA 序列分析、PCR 和原位杂交检测突变。另外，突变有可能影响蛋白的表达，因此用 Northern 印迹法、Western 印迹法可间接判断基因有无突变。

本发明的序列对染色体鉴定也是有价值的。该序列会特异性地针对某条人染色体具体位置并且可以与其杂交。目前，需要鉴定染色体上的各基因的具体位点。现在，只有很少的基于实际序列数据 (重复多态性) 的染色体标记物可用于标记染色体位置。根据本发明，为了将这些序列与疾病相关基因相关联，其重要的第一步就是将这些 DNA 序列定位于染色体上。

简而言之，根据 cDNA 制备 PCR 引物 (优选 15-35bp)，可以将序列定位于染色体上。然后，将这些引物用于 PCR 筛选含各条人染色体的体细胞杂合细胞。只有那些含有相应于引物的人基因的杂合细胞会产生扩增的片段。

体细胞杂合细胞的 PCR 定位法，是将 DNA 定位到具体染色体的快捷方法。使用本发明的寡核苷酸引物，通过类似方法，可利用一组来自特定染色体的片段或大量基因组克隆而实现亚定位。可用于染色体定位的其它类似策略包括原位杂交、用标记的流式分选的染色体预筛选和杂交预选，从而构建染色体特异的 cDNA 库。

将cDNA克隆与中期染色体进行荧光原位杂交(FISH)，可以在一个步骤中精确地进行染色体定位。此技术的综述参见Verma等, Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, New York(1988)。

一旦序列被定位到准确的染色体位置，此序列在染色体上的物理位置就可以与基因图数据相关联。这些数据可见于V. McKusick, Mendelian Inheritance in Man(可通过与 Johns Hopkins University Welch Medical Library 联机获得)。然后可通过连锁分析，确定基因与业已定位到染色体区域上的疾病之间的关系。

接着，需要测定患病和未患病个体间的cDNA或基因组序列差异。如果在一些或所有的患病个体中观察到某突变，而该突变在任何正常个体中未观察到，则该突变可能是疾病的病因。比较患病和未患病个体，通常涉及首先寻找染色体中结构的变化，如从染色体水平可见的或用基于cDNA序列的PCR可检测的缺失或易位。根据目前的物理作图和基因定位技术的分辨能力，被精确定位至与疾病有关的染色体区域的cDNA，可以是50至500个潜在致病基因间之一种(假定1兆碱基作图分辨能力和每20kb对应于一个基因)。

可以将本发明的多肽、多核苷酸及其模拟物、激动剂、拮抗剂和抑制剂与合适的药物载体组合后使用。这些载体可以是水、葡萄糖、乙醇、盐类、缓冲液、甘油以及它们的组合。组合物包含安全有效量的多肽或拮抗剂以及不影响药物效果的载体和赋形剂。这些组合物可以作为药物用于疾病治疗。

本发明还提供含有一种或多种容器的药盒或试剂盒，容器中装有一种或多种本发明的药用组合物成分。与这些容器一起，可以有由制造、使用或销售药品或生物制品的政府管理机构所给出的指示性提示，该提示反映出生产、使用或销售的政府管理机构许可其在人体上施用。此外，本发明的多肽可以与其它的治疗化合物结合使用。

药物组合物可以以方便的方式给药，如通过局部、静脉内、腹膜内、肌肉、皮下、鼻内或皮内的给药途径。N-乙酰半乳糖胺转移酶-28 以有效地治疗和/或预防具体的适应症的量来给药。施用于患者的 N-乙酰半乳糖胺转移酶-28 的量和剂量范围将取决于许多因素，如给药方式、待治疗者的健康条件和诊断医生的判断。

实施例

下面结合具体实施例，进一步阐述本发明。应理解，这些实施例仅用于说

明本发明而不适用于限制本发明的范围。下列实施例中未注明具体条件的实验方法，通常按照常规条件如 Sambrook 等人，分子克隆：实验室手册 (New York: Cold Spring Harbor Laboratory Press, 1989) 中所述的条件，或按照制造厂商所建议的条件。

实施例 1 N-乙酰半乳糖胺转移酶-28 的克隆

用异硫氰酸胍/酚/氯仿一步法提取人胎脑总RNA。用Quik mRNA Isolation Kit (Qiegene 公司产品) 从总RNA中分离poly(A) mRNA。2ug poly(A) mRNA经逆转录形成cDNA。用Smart cDNA克隆试剂盒(购自Clontech)将cDNA片段定向插入到pBSK(+) 载体(Clontech公司产品)的多克隆位点上，转化DH5 α ，细菌形成cDNA文库。用Dye terminate cycle reaction sequencing kit (Perkin-Elmer公司产品) 和ABI 377 自动测序仪(Perkin-Elmer公司)测定所有克隆的5'和3'末端的序列。将测定的cDNA序列与已有的公共DNA序列数据库 (Genbank) 进行比较，结果发现其中一个克隆2105d02的cDNA序列为新的DNA。通过合成一系列引物对该克隆所含的插入cDNA片段进行双向测定。结果表明，2105d02克隆所含的全长cDNA为1271bp (如Seq ID NO: 1所示)，从第193bp至966bp有一个 774bp的开放阅读框架 (ORF)，编码一个新的蛋白质 (如Seq ID NO: 2所示)。我们将此克隆命名为pBS-2105d02，编码的蛋白质命名为N-乙酰半乳糖胺转移酶-28。

实施例2 cDNA 克隆的同源检索

将本发明的N-乙酰半乳糖胺转移酶-28的序列及其编码的蛋白序列，用Blast 程序 (Basic local alignment search tool) [Altschul, SF et al. J. Mol. Biol. 1990; 215: 403-10]，在Genbank、Swissport等数据库进行同源检索。与本发明的N-乙酰半乳糖胺转移酶-28同源性最高的基因是一种已知的N-乙酰半乳糖胺转移酶，其编码的蛋白在Genbank的准入号为U73820。蛋白质同源结果示于图1，两者高度同源，其相同性为34%；相似性为58%。

实施例3 用RT-PCR方法克隆编码N-乙酰半乳糖胺转移酶-28的基因

用胎脑细胞总RNA为模板，以oligo-dT为引物进行逆转录反应合成cDNA，用Qiagene的试剂盒纯化后，用下列引物进行PCR扩增：

Primer1: 5'- AACTTTGAGATAGAAGAGTACCCG -3' (SEQ ID NO: 3)

Primer2: 5'- CATAGGCCGAGGCCGCCGACATGT -3' (SEQ ID NO: 4)

Primer1为位于SEQ ID NO: 1的5'端的第1bp开始的正向序列；

Primer2为SEQ ID NO:1的中的3'端反向序列。

扩增反应的条件: 在50 μ l的反应体积中含有50mmol/L KCl, 10mmol/L Tris-HCl pH8.5, 1.5mmol/L MgCl₂, 200 μ mol/L dNTP, 10pmol引物, 1U的Taq DNA聚合酶(Clontech公司产品)。在PE9600型DNA热循环仪(Perkin-Elmer公司)上按下列条件反应25个周期: 94°C 30sec; 55°C 30sec; 72°C 2min。在RT-PCR时同时设 β -actin为阳性对照和模板空白为阴性对照。扩增产物用QIAGEN公司的试剂盒纯化, 用TA克隆试剂盒连接到pCR载体上(Invitrogen公司产品)。DNA序列分析结果表明PCR产物的DNA序列与SEQ ID NO:1所示的1-1271bp完全相同。

实施例4 Northern 印迹法分析N-乙酰半乳糖胺转移酶-28基因的表达

用一步法提取总RNA[Anal. Biochem 1987, 162, 156-159]。该法包括酸性硫氰酸胍苯酚-氯仿抽提。即用4M异硫氰酸胍-25mM柠檬酸钠, 0.2M乙酸钠(pH4.0)对组织进行匀浆, 加入1倍体积的苯酚和1/5体积的氯仿-异戊醇(49:1), 混合后离心。吸出水相层, 加入异丙醇(0.8体积)并将混合物离心得到RNA沉淀。将得到的RNA沉淀用70%乙醇洗涤, 干燥并溶于水中。用20 μ g RNA, 在含20mM 3-(N-吗啉代)丙磺酸(pH7.0)-5mM乙酸钠-1mM EDTA-2.2M甲醛的1.2%琼脂糖凝胶上进行电泳。然后转移至硝酸纤维素膜上。用 α -³²P dATP通过随机引物法制备³²P-标记的DNA探针。所用的DNA探针为图1所示的PCR扩增的N-乙酰半乳糖胺转移酶-28编码区序列(193bp至966bp)。将³²P-标记的探针(约2 \times 10⁶cpm/ml)与转移了RNA的硝酸纤维素膜在一溶液于42°C杂交过夜, 该溶液包含50%甲酰胺-25mM KH₂PO₄(pH7.4)-5 \times SSC-5 \times Denhardt's溶液和200 μ g/ml鲑精DNA。杂交之后, 将滤膜在1 \times SSC-0.1%SDS中于55°C洗30min。然后, 用Phosphor Imager进行分析和定量。

实施例5 重组N-乙酰半乳糖胺转移酶-28的体外表达、分离和纯化

根据SEQ ID NO:1和图1所示的编码区序列, 设计出一对特异性扩增引物, 序列如下:

Primer3: 5'-CCCCATATGATGGAAGTCTACGGGGGCGAGAAT-3' (Seq ID No:5)

Primer4: 5'-CATGGATCCTCAGGACGCGAGGCTCCTCAGGAC-3' (Seq ID No:6)

此两段引物的5'端分别含有NdeI和BamHI酶切位点, 其后分别为目的基因5'端和3'端的编码序列, NdeI和BamHI酶切位点相应于表达载体质粒pET-28b(+)(Novagen公司产品, Cat.No. 69865.3)上的选择性内切酶位点。以含有全长目的基因的pBS-2105d02质粒为模板, 进行PCR反应。PCR反应条件为: 总体积50 μ l

中含pBS-2105d02质粒10pg、引物Primer-3和Primer-4分别为10pmol、Advantage polymerase Mix (Clontech公司产品) 1 μ l。循环参数: 94°C 20s, 60°C 30s, 68°C 2 min, 共25个循环。用NdeI和BamHI分别对扩增产物和质粒pET-28(+)进行双酶切, 分别回收大片段, 并用T4连接酶连接。连接产物转化用氯化钙法大肠杆菌DH5 α , 在含卡那霉素(终浓度30 μ g/ml)的LB平板培养过夜后, 用菌落PCR方法筛选阳性克隆, 并进行测序。挑选序列正确的阳性克隆(pET-2105d02)用氯化钙法将重组质粒转化大肠杆菌BL21(DE3)plySs(Novagen公司产品)。在含卡那霉素(终浓度30 μ g/ml)的LB液体培养基中, 宿主菌BL21(pET-2105d02)在37°C培养至对数生长期, 加入IPTG至终浓度1mmol/L, 继续培养5小时。离心收集菌体, 经超声波破菌, 离心收集上清, 用能与6个组氨酸(6His-Tag)结合的亲和层析柱His. Bind Quick Cartridge (Novagen公司产品)进行层析, 得到了纯化的目的蛋白N-乙酰半乳糖胺转移酶-28。经SDS-PAGE电泳, 在28kDa处得到一单一的条带(图2)。将该条带转移至PVDF膜上用Edams水解法进行N-端氨基酸序列分析, 结果N-端15个氨基酸与SEQ ID NO: 2所示的N-端15个氨基酸残基完全相同。

实施例6 抗N-乙酰半乳糖胺转移酶-28抗体的产生

用多肽合成仪(PE公司产品)合成下述N-乙酰半乳糖胺转移酶-28特异性的多肽:

NH₂-Met-Glu-Val-Tyr-Gly-Gly-Glu-Asn-Val-Glu-Leu-Gly-Ile-Arg-Val-COOH

(SEQ ID NO: 7)。将该多肽分别与血蓝蛋白和牛血清白蛋白耦合形成复合物, 方法参见: Avrameas, et al. Immunochimistry, 1969; 6: 43。用4mg上述血蓝蛋白多肽复合物加上完全弗氏佐剂免疫家兔, 15天后再用血蓝蛋白多肽复合物加不完全弗氏佐剂加强免疫一次。采用经15 μ g/ml牛血清白蛋白多肽复合物包被的滴定板做ELISA测定兔血清中抗体的滴度。用蛋白A-Sepharose从抗体阳性的家兔血清中分离总IgG。将多肽结合于溴化氰活化的Sepharose4B柱上, 用亲和层析法从总IgG中分离抗多肽抗体。免疫沉淀法证明纯化的抗体可特异性地与N-乙酰半乳糖胺转移酶-28结合。

实施例7 本发明的多核苷酸片段用作杂交探针的应用

从本发明的多核苷酸中挑选出合适的寡核苷酸片段用作杂交探针有多方面的用途, 如用该探针可与不同来源的正常组织或病理组织的基因组或cDNA文库杂交以鉴定其是否含有本发明的多核苷酸序列和检出同源的多核苷酸序列, 进一步还可

用该探针检测本发明的多核苷酸序列或其同源的多核苷酸序列在正常组织或病理组织细胞中的表达是否异常。

本实施例的目的是从本发明的多核苷酸 SEQ ID NO: 1 中挑选出合适的寡核苷酸片段用作杂交探针，并用滤膜杂交方法鉴定一些组织中是否含有本发明的多核苷酸序列或其同源的多核苷酸序列。滤膜杂交方法包括斑点印迹法、Southern 印迹法、Northern 印迹法和复印方法等，它们都是将待测的多核苷酸样品固定在滤膜上后使用基本相同的步骤杂交。这些相同的步骤是：固定了样品的滤膜首先用不含探针的杂交缓冲液进行预杂交，以使滤膜上样品的非特异性的结合部位被载体和合成的多聚物所饱和，然后预杂交液被含有标记探针的杂交缓冲液替换，并保温使探针与靶核酸杂交。杂交步骤之后，未杂交上的探针被一系列洗膜步骤除掉。本实施例利用较高强度的洗膜条件（如较低盐浓度和较高的温度），以使杂交背景降低且只保留特异性强的信号。本实施例选用的探针包括两类：第一类探针是完全与本发明的多核苷酸 SEQ ID NO: 1 相同或互补的寡核苷酸片段；第二类探针是部分与本发明的多核苷酸 SEQ ID NO: 1 相同或互补的寡核苷酸片段。本实施例选用斑点印迹法将样品固定在滤膜上，在较高强度的洗膜条件下，第一类探针与样品的杂交特异性最强而得以保留。

一、 探针的选用

从本发明的多核苷酸 SEQ ID NO: 1 中选择寡核苷酸片段用作杂交探针，应遵循以下原则和需要考虑的几个方面：

- 1, 探针大小优选范围为 18-50 个核苷酸；
- 2, GC 含量为 30%-70%，超过则非特异性杂交增加；
- 3, 探针内部应无互补区域；
- 4, 符合以上条件的可作为初选探针，然后进一步作计算机序列分析，包括将该初选探针分别与其来源序列区域（即 SEQ ID NO: 1）和其它已知的基因组序列及其互补区进行同源性比较，若与非靶分子区域的同源性大于 85% 或者有超过 15 个连续碱基完全相同，则该初选探针一般就不应该使用；

5, 初选探针是否最终选定为有实际应用价值的探针还应进一步由实验确定。

完成以上各方面的分析后挑选并合成以下二个探针：

探针 1 (probe1), 属于第一类探针，与 SEQ ID NO: 1 的基因片段完全同源或互补 (41nt)：

5'-TCGAAGTCTACGGGGGCCGAGAATGTGGAGCTTGGGATCAGG-3' (SEQ ID NO: 8)

探针 2 (probe2), 属于第二类探针，相当于 SEQ ID NO: 1 的基因片段或其

互补片段的替换突变序列 (41nt):

5'-TGGAAGTCTACGGGGCGAGCATGTGGAGCTTGGGATCAGG-3' (SEQ ID NO: 9)

与以下具体实验步骤有关的其它未列出的常用试剂及其配制方法请参考文献: DNA PROBES G. H. Keller; M. M. Manak; Stockton Press, 1989 (USA) 以及更常用的分子克隆实验手册书籍如《分子克隆实验指南》(1998 年第二版) [美] 萨姆布鲁克等著, 科学出版社。

样品制备:

1, 从新鲜或冰冻组织中提取 DNA

步骤: 1) 将新鲜或新鲜解冻的正常肝组织放入浸在冰上并盛有磷酸盐缓冲液 (PBS) 的平皿中。用剪刀或手术刀将组织切成小块。操作中应保持组织湿润。2) 以 1000g 离心切碎组织 10 分钟。3) 用冷匀浆缓冲液 (0.25mol/L 蔗糖; 25mmol/L Tris-HCl, pH7.5; 25mmol/L NaCl; 25mmol/L MgCl₂) 悬浮沉淀 (大约 10ml/g)。4) 在 4°C 用电动匀浆器以全速匀浆组织悬液, 直至组织被完全破碎。5) 1000g 离心 10 分钟。6) 用重悬细胞沉淀 (每 0.1g 最初组织样品加 1-5ml), 再以 1000g 离心 10 分钟。7) 用裂解缓冲液重悬沉淀 (每 0.1g 最初组织样品加 1ml), 然后接以下的苯酚抽提法。

2, DNA 的苯酚抽提法

步骤: 1) 用 1-10ml 冷 PBS 洗细胞, 1000g 离心 10 分钟。2) 用冷细胞裂解液重悬浮沉淀的细胞 (1×10^8 细胞/ml) 最少应用 100ul 裂解缓冲液。3) 加 SDS 至终浓度为 1%, 如果在重悬细胞之前将 SDS 直接加入到细胞沉淀中, 细胞可能会形成大的团块而难以破碎, 并降低总产率。这一点在抽提 $>10^7$ 细胞时特别严重。4) 加蛋白酶 K 至终浓度 200ug/ml。5) 50°C 保温反应 1 小时或在 37°C 轻轻振摇过夜。6) 用等体积苯酚: 氯仿: 异戊醇 (25: 24: 1) 抽提, 在小离心机管中离心 10 分钟。两相应清楚分离, 否则重新进行离心。7) 将水相转移至新管。8) 用等体积氯仿: 异戊醇 (24: 1) 抽提, 离心 10 分钟。9) 将含 DNA 的水相转移至新管。然后进行 DNA 的纯化和乙醇沉淀。

3, DNA 的纯化和乙醇沉淀

步骤: 1) 将 1/10 体积 2mol/L 醋酸钠和 2 倍体积冷 100%乙醇加到 DNA 溶液中, 混匀。在 -20°C 放置 1 小时或过夜。2) 离心 10 分钟。3) 小心吸出或倒出乙醇。4) 用 70%冷乙醇 500ul 洗涤沉淀, 离心 5 分钟。5) 小心吸出或倒出乙醇。用 500ul 冷乙醇洗涤沉淀, 离心 5 分钟。6) 小心吸出或倒出乙醇, 然后在吸水纸上倒置使残余乙醇流尽。空气干燥 10-15 分钟, 以使表面乙醇挥发。注意不要使沉

淀完全干燥，否则较难重新溶解。7) 以小体积 TE 或水重悬 DNA 沉淀。低速涡旋振荡或用滴管吹吸，同时逐渐增加 TE，混合至 DNA 充分溶解，每 $1-5 \times 10^6$ 细胞所提取的大约加 1 μ l。

以下第 8-13 步骤仅用于必须除去污染时，否则可直接进行第 14 步骤。

8) 将 RNA 酶 A 加到 DNA 溶液中，终浓度为 100 μ g/ml，37°C 保温 30 分钟。9) 加入 SDS 和蛋白酶 K，终浓度分别为 0.5% 和 100 μ g/ml。37°C 保温 30 分钟。10) 用等体积的苯酚：氯仿：异戊醇 (25: 24: 1) 抽提反应液，离心 10 分钟。11) 小心移出水相，用等体积的氯仿：异戊醇 (24: 1) 重新抽提，离心 10 分钟。12) 小心移出水相，加 1/10 体积 2mol/L 醋酸钠和 2.5 体积冷乙醇，混匀置 -20°C 1 小时。13) 用 70% 乙醇及 100% 乙醇洗涤沉淀，空气干燥，重悬核酸，过程同第 3-6 步骤。14) 测定 A_{260} 和 A_{280} 以检测 DNA 的纯度及产率。15) 分装后存放于 -20°C。

样膜的制备：

1) 取 4×2 张适当大小的硝酸纤维素膜 (NC 膜)，用铅笔在其上轻轻标出点样位置及样号，每一探针需两张 NC 膜，以便在后面的实验步骤中分别用高强度条件和强度条件洗膜。

2) 吸取及对照各 15 微升，点于样膜上，在室温中晾干。

3) 置于浸润有 0.1mol/L NaOH, 1.5mol/L NaCl 的滤纸上 5 分钟 (两次)，晾干置于浸润有 0.5mol/L Tris-HCl (pH7.0), 3mol/L NaCl 的滤纸上 5 分钟 (两次)，晾干。

4) 夹于干净滤纸中，以铝箔包好，60-80°C 真空干燥 2 小时。

探针的标记

1) 3 μ l Probe (0.10D/10 μ l)，加入 2 μ l Kinase 缓冲液，8-10 uCi γ - 32 P-dATP+2U Kinase，以补加至终体积 20 μ l。

2) 37°C 保温 2 小时。

3) 加 1/5 体积的溴酚蓝指示剂 (BPB)。

4) 过 Sephadex G-50 柱。

5) 至有 32 P-Probe 洗出前开始收集第一峰 (可用 Monitor 监测)。

6) 5 滴/管，收集 10-15 管。

7) 用液体闪烁仪监测同位素量。

8) 合并第一峰的收集液后即所需制备的 32 P-Probe (第二峰为游离 γ - 32 P-dATP)。

预杂交

将样膜置于塑料袋中,加入 3-10mg 预杂交液(10×Denhardt's; 6×SSC, 0.1mg/ml CT DNA (小牛胸腺 DNA)). 封好袋口后, 68°C 水浴摇 2 小时。

杂交

将塑料袋剪去一角, 加入制备好的探针, 封好袋口后, 42°C 水浴摇过夜。

洗膜:

高强度洗膜:

- 1) 取出已杂交好的样膜。
- 2) 2×SSC, 0.1%SDS 中, 40°C 洗 15 分钟 (2 次)。
- 3) 0.1×SSC, 0.1%SDS 中, 40°C 洗 15 分钟 (2 次)。
- 4) 0.1×SSC, 0.1%SDS 中, 55°C 洗 30 分钟 (2 次), 室温晾干。

低强度洗膜:

- 1) 取出已杂交好的样膜。
- 2) 2×SSC, 0.1%SDS 中, 37°C 洗 15 分钟 (2 次)。
- 3) 0.1×SSC, 0.1%SDS 中, 37°C 洗 15 分钟 (2 次)。
- 4) 0.1×SSC, 0.1%SDS 中, 40°C 洗 15 分钟 (2 次), 室温晾干。

X-光自显影:

-70°C, X-光自显影 (压片时间根据杂交斑放射性强弱而定)。

实验结果:

采用低强度洗膜条件所进行的杂交实验, 以上两个探针杂交斑放射性强弱没有明显区别; 而采用高强度洗膜条件所进行的杂交实验, 探针 1 的杂交斑放射性强于另一个探针杂交斑的放射性强。因而可用探针 1 定性和定量地分析本发明的多核苷酸在不同组织中的存在和差异表达。

权利要求

1、一种分离的多肽 - N-乙酰半乳糖胺转移酶-28，其特征在于它包含有：SEQ ID NO: 2 所示的氨基酸序列的多肽、或其多肽的活性片段、类似物或衍生物。

2、如权利要求 1 所述的多肽，其特征在于所述多肽、类似物或衍生物的氨基酸序列具有与 SEQ ID NO: 2 所示的氨基酸序列至少 95%的相同性。

3、如权利要求 2 所述的多肽，其特征在于它包含具有 SEQ ID NO: 2 所示的氨基酸序列的多肽。

4、一种分离的多核苷酸，其特征在于所述多核苷酸包含选自下组中的一种：

(a) 编码具有 SEQ ID NO: 2 所示氨基酸序列的多肽或其片段、类似物、衍生物的多核苷酸；

(b) 与多核苷酸 (a) 互补的多核苷酸；或

(c) 与 (a) 或 (b) 有至少 70%相同性的多核苷酸。

5、如权利要求 4 所述的多核苷酸，其特征在于所述多核苷酸包含编码具有 SEQ ID NO: 2 所示氨基酸序列的多核苷酸。

6、如权利要求 4 所述的多核苷酸，其特征在于所述多核苷酸的序列包含有 SEQ ID NO: 1 中 193-966 位的序列或 SEQ ID NO: 1 中 1-1271 位的序列。

7、一种含有外源多核苷酸的重组载体，其特征在于它是由权利要求 4-6 中的任一权利要求所述多核苷酸与质粒、病毒或运载体表达载体构建而成的重组载体。

8、一种含有外源多核苷酸的遗传工程化宿主细胞，其特征在于它是选自于下列一种宿主细胞：

(a) 用权利要求 7 所述的重组载体转化或转导的宿主细胞；或

(b) 用权利要求 4-6 中的任一权利要求所述多核苷酸转化或转导的宿主细胞。

9、一种具有 N-乙酰半乳糖胺转移酶-28 活性的多肽的制备方法，其特征在于所述方法包括：

(a) 在表达 N-乙酰半乳糖胺转移酶-28 条件下，培养权利要求 8 所述的工程化宿主细胞；

(b) 从培养物中分离出具有 N-乙酰半乳糖胺转移酶-28 活性的多肽。

10、一种能与多肽结合的抗体,其特征在於所述抗体是能与 N-乙酰半乳糖胺转移酶-28 特异性结合的抗体。

11、一类模拟或调节多肽活性或表达的化合物,其特征在於它们是模拟、促进、拮抗或抑制 N-乙酰半乳糖胺转移酶-28 的活性的化合物。

12、如权利要求 11 所述的化合物,其特征在於它是 SEQ ID NO:1 所示的多核苷酸序列或其片段的反义序列。

13、一种权利要求 11 所述化合物的应用,其特征在於所述化合物用于调节 N-乙酰半乳糖胺转移酶-28 在体内、体外活性的方法。

14、一种检测与权利要求 1-3 中的任一权利要求所述多肽相关的疾病或疾病易感性的方法,其特征在於其包括检测所述多肽的表达量,或者检测所述多肽的活性,或者检测多核苷酸中引起所述多肽表达量或活性异常的核苷酸变异。

15、如权利要求 1-3 中的任一权利要求所述多肽的应用,其特征在於它应用于筛选 N-乙酰半乳糖胺转移酶-28 的模拟物、激动剂,拮抗剂或抑制剂;或者用于肽指纹图谱鉴定。

16、如权利要求 4-6 中的任一权利要求所述的核酸分子的应用,其特征在於它作为引物用于核酸扩增反应,或者作为探针用于杂交反应,或者用于制造基因芯片或微阵列。

17、如权利要求 1-6 及 11 中的任一权利要求所述的多肽、多核苷酸或化合物的应用,其特征在於用所述多肽、多核苷酸或其模拟物、激动剂、拮抗剂或抑制剂以安全有效剂量与药学上可接受的载体组成作为诊断或治疗与 N-乙酰半乳糖胺转移酶-28 异常相关的疾病的药物组合物。

18、权利要求 1-6 及 11 中的任一权利要求所述的多肽、多核苷酸或化合物的应用,其特征在於用所述多肽、多核苷酸或化合物制备用于治疗如恶性肿瘤,血液病, HIV 感染和免疫性疾病和各类炎症的药物。

相同性= 87/251 (34%), 相似性= 147/251 (58%)

Query: 1 MEVYGGENVELGIRVWQCGGSVEVLPCSRIAHIERAHKPYT-EDLTAHV-RRNALRVAEV 58
M+++GGEN+E+ R+WQCGG++E++ CS + H+ R PYT T + +N R+AEV

Sbjct: 313 MDIWGGENLEISFRIWQCGGTLEIVTCSHVGHVFRKATPYTFPGGTGQIINKNNRRLAEV 372

Query: 59 WMDEFKSHVYMAWNIPQEDSGIDIGDITARKALRKQLQCKTFRWYLVSVYPEMRMYSDII 118
WMDEFK+ Y+ I + +D GDI++R LR++LQCK F WYL ++YP+ ++

Sbjct: 373 WMDEFKNFFYI---ISPGVTKVDYGDISSRLGLARKLQCKPFSWYLENIYPDSQIPRHYF 429

Query: 119 AYGVLQNSLKTDLCLDQGPDTENVPI-MYICHGMTQPNVY-YTSSQQIHVGILSPTVDDD 176
+ G ++N ++T+ CLD EN + ++ CHGM V+ YT++++I

Sbjct: 430 SLGEIRN-VETNQCLDNMARKENEKVGIFNCHGMGCNQVFSY TANKEIRT----- 478

Query: 177 DNRCLVDV---NSRPLIECSYAKAKRMKLRWQFSQGG-PIQNRKSKRCLELQBNSDLEF 232
D+ CL DV N +++C + K ++ W++ +Q+ S +CL+ D +

Sbjct: 479 DDLCL-DVSKLNGPVTMLKCHHLKGNQL---WEYDPVKLTLQHVNSNQCLDKATEEDSQV 534

Query: 233 GFQLVLQKCSG---QHWSITNV 251
++ C+G Q W + NV

Sbjct: 535 PS---IRDCTGSRSQWLLRNV 553

Query 表示 N-乙酰半乳糖胺转移酶-28

Sbjct 表示 N-乙酰半乳糖胺转移酶

图 1

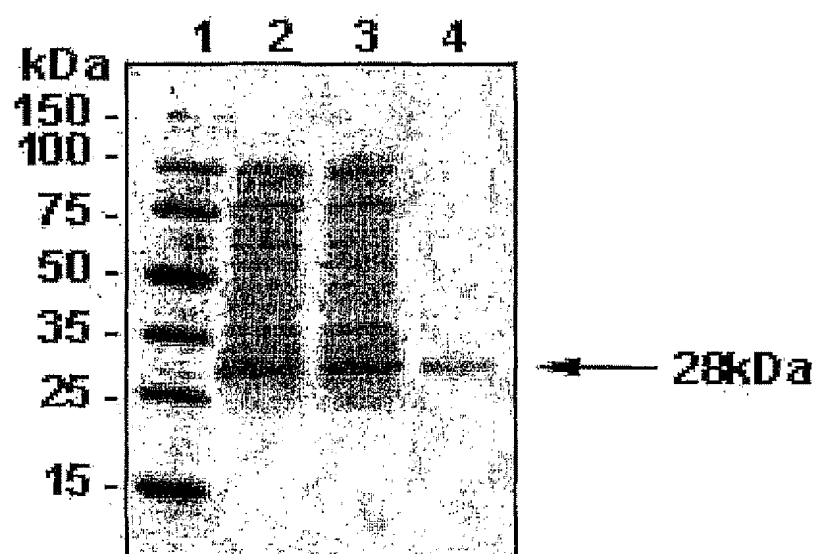


图 2

序 列 表

(1) 一般信息

(ii) 发明名称: N-乙酰半乳糖胺转移酶-28及其编码序列

(iii) 序列数目: 9

(2) SEQ ID NO: 1的信息

(i) 序列特征

(A) 长度: 1271bp

(B) 类型: 核酸

(C) 链性: 双链

(D) 拓扑结构: 线性

(ii) 分子类型: cDNA

(xi) 序列描述: SEQ ID NO: 1:

```
1  AACTTTGAGATAGAAGAGTACCCGCTGGCTGCCAGGGCTTTGACTGGGAGCTGTGCTGC
61  CGCTACCTAAATCCCCCAAGGCCTGGTGAAGCTGGAGAACTCCACAGCGCCAATCAGG
121 AGCCCTGCCCACATTGGCTGCTTCATTGTGGACCGGCAGTACTTCCAGGAGATCGGCCTG
181 CTGGACGAAGGCATGGAAGTCTACGGGGGCGAGAATGTGGAGCTTGGGATCAGGGTGTGG
241 CAGTGTGGCGGGAGTGTGGAGGTCCTGCCCTGCTCACGGATTGCCACATTGAGCGAGCC
301 CACAAGCCCTACACAGAGGACCTACCGCCCATGTCCCGAGGAACGCTCTCAGGGTGGCT
361 GAAGTCTGGATGGATGAATTTAAAAGCCACGTCTACATGGCATGGAACATAACCGCAGGAG
421 GACTCAGGAATTGACATTGGGGACATCACTGCAAGGAAGGCTCTCAGGAAACAGCTGCAG
481 TGCAAGACCTTCCGGTGGTACCTGGTCAGCGTGTACCCAGAGATGAGGATGTACTCCGAC
541 ATCATTGCCTATGGAGTGCTGCAGAATTCTCTGAAGACTGATTTGTGTCTTGACCAGGGG
601 CCAGATACAGAGAATGTCCCCATCATGTACATCTGCCATGGGATGACGCCTCAGAACGTG
661 TACTACACGAGCAGTCAGCAGATCCATGTGGGCATTCTGAGCCCCACCGTGGATGATGAT
721 GACAACCGATGCCTGGTGGACGTCAACAGCCGGCCCCGGCTCATCGAATGCAGCTACGCC
781 AAAGCCAAGAGGATGAAGCTTCGCTGGCAGTTCTCTCAGGGAGGACCCATCCAGAACCGC
841 AAGTCTAAGCGCTGTCTGGAGCTGCAGGAGAATAGCGACCTGGAGTTCGGCTTCCAGCTG
901 GTGTTGCAGAAGTGCTCGGGCCAGCACTGGAGCATCACCAACGTCCTGAGGAGCCTCGCG
961 TCCTGACCCACCGGGGCCACTTCCGGCTGCCTCTTTGCTACTGTGTAGCACCTGCTGCAA
1021 CGTTGCCTGCTGTCCACGTGGGGTTGTTGGAGTCTGGGGAACCAGGTTAGTGGGCCCCC
```

1081 AAGAAGAGCTTTTTATTTCTATTCAATTTTCATGGAGTTTATAGAAAAGATGCTGATTGG
 1141 TAGGTGATGGTATGATATCAAACATTTTGCAGTTGTAAATAGGGGACAGATGGAAAATA
 1201 TTTATAACTGACAATAAAATATTATTAAGAAAAGGAAAAAAAAAAAAACATGTCGGCCGC
 1261 CTCGGCCTATG

(3) SEQ ID NO: 2的信息

(i) 序列特征

(A) 长度: 257个氨基酸

(B) 类型: 氨基酸

(D) 拓扑结构: 线性

(ii) 分子类型: 多肽

(xi) 序列描述: SEQ ID NO: 2:

1	Met	Glu	Val	Tyr	Gly	Gly	Glu	Asn	Val	Glu	Leu	Gly	Ile	Arg	Val
16	Trp	Gln	Cys	Gly	Gly	Ser	Val	Glu	Val	Leu	Pro	Cys	Ser	Arg	Ile
31	Ala	His	Ile	Glu	Arg	Ala	His	Lys	Pro	Tyr	Thr	Glu	Asp	Leu	Thr
46	Ala	His	Val	Arg	Arg	Asn	Ala	Leu	Arg	Val	Ala	Glu	Val	Trp	Met
61	Asp	Glu	Phe	Lys	Ser	His	Val	Tyr	Met	Ala	Trp	Asn	Ile	Pro	Gln
76	Glu	Asp	Ser	Gly	Ile	Asp	Ile	Gly	Asp	Ile	Thr	Ala	Arg	Lys	Ala
91	Leu	Arg	Lys	Gln	Leu	Gln	Cys	Lys	Thr	Phe	Arg	Trp	Tyr	Leu	Val
106	Ser	Val	Tyr	Pro	Glu	Met	Arg	Met	Tyr	Ser	Asp	Ile	Ile	Ala	Tyr
121	Gly	Val	Leu	Gln	Asn	Ser	Leu	Lys	Thr	Asp	Leu	Cys	Leu	Asp	Gln
136	Gly	Pro	Asp	Thr	Glu	Asn	Val	Pro	Ile	Met	Tyr	Ile	Cys	His	Gly
151	Met	Thr	Pro	Gln	Asn	Val	Tyr	Tyr	Thr	Ser	Ser	Gln	Gln	Ile	His
166	Val	Gly	Ile	Leu	Ser	Pro	Thr	Val	Asp	Asp	Asp	Asp	Asn	Arg	Cys
181	Leu	Val	Asp	Val	Asn	Ser	Arg	Pro	Arg	Leu	Ile	Glu	Cys	Ser	Tyr
196	Ala	Lys	Ala	Lys	Arg	Met	Lys	Leu	Arg	Trp	Gln	Phe	Ser	Gln	Gly
211	Gly	Pro	Ile	Gln	Asn	Arg	Lys	Ser	Lys	Arg	Cys	Leu	Glu	Leu	Gln
226	Glu	Asn	Ser	Asp	Leu	Glu	Phe	Gly	Phe	Gln	Leu	Val	Leu	Gln	Lys
241	Cys	Ser	Gly	Gln	His	Trp	Ser	Ile	Thr	Asn	Val	Leu	Arg	Ser	Leu
256	Ala	Ser													

(4) SEQ ID NO: 3的信息

(i) 序列特征

- (A) 长度: 24碱基
- (B) 类型: 核酸
- (C) 链性: 单链
- (D) 拓扑结构: 线性
- (ii) 分子类型: 寡核苷酸
- (xi) 序列描述: SEQ ID NO: 3:

AACTTTGAGATAGAAGAGTACCCG

24

(5) SEQ ID NO: 4的信息

- (i) 序列特征
 - (A) 长度: 24碱基
 - (B) 类型: 核酸
 - (C) 链性: 单链
 - (D) 拓扑结构: 线性
- (ii) 分子类型: 寡核苷酸
- (xi) 序列描述: SEQ ID NO: 4:

CATAGGCCGAGGCGGCCGACATGT

24

(6) SEQ ID NO: 5的信息

- (i) 序列特征
 - (A) 长度: 33碱基
 - (B) 类型: 核酸
 - (C) 链性: 单链
 - (D) 拓扑结构: 线性
- (ii) 分子类型: 寡核苷酸
- (xi) 序列描述: SEQ ID NO : 5:

CCCCATATGATGGAAGTCTACGGGGGCGAGAAT

33

(7) SEQ ID NO: 6的信息

- (i) 序列特征
 - (A) 长度: 33碱基
 - (B) 类型: 核酸
 - (C) 链性: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

(xi) 序列描述: SEQ ID NO : 6:

CATGGATCCTCAGGACGCGAGGCTCCTCAGGAC

33

(8) SEQ ID NO: 7 的信息

(i) 序列特征

(A) 长度: 15个氨基酸

(B) 类型: 氨基酸

(D) 拓扑结构: 线性

(ii) 分子类型: 多肽

(xi) 序列描述: SEQ ID NO: 7:

Met-Glu-Val-Tyr-Gly-Gly-Glu-Asn-Val-Glu-Leu-Gly-Ile-Arg-Val 15

(9) SEQ ID NO: 8 的信息

(i) 序列特征

(A) 长度: 41碱基

(B) 类型: 核酸

(C) 链性: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

(xi) 序列描述: SEQ ID NO : 8:

TGGAAGTCTACGGGGCGAGAATGTGGAGCTTGGGATCAGG

41

(10) SEQ ID NO: 9 的信息

(i) 序列特征

(A) 长度: 41碱基

(B) 类型: 核酸

(C) 链性: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

(xi) 序列描述: SEQ ID NO : 9:

TGGAAGTCTACGGGGCGAGCATGTGGAGCTTGGGATCAGG

41

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN01/00749

A. CLASSIFICATION OF SUBJECT MATTER

IPC⁷: C12N15/54, C12N9/10, C07K14/435, A61K38/45

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched(classification system followed by classification symbols)

IPC⁷: C12N15/54, C12N9/10, C07K14/435, A61K38/45

Documentation searched other than minimum documentation to the extent that such documents are included in the field searched

Electronic data base consulted during the international search(name of data base and, where practicable, search terms used)

WPI, GenBank+EMBL+DDBJ+Swiss-Prot+PIR+PDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant claim No.
A	EMBL AJ271385 HSA271385 2000-04-26 See the sequence	1-18
A	GenBank AC005833 1999-01-30 See the sequence	1-18
A	EMBL AL022577 HS353H6 1999-11-23 See the sequence	1-18
A	GenBank U91318 HUU91318 2000-01-10 See the sequence	1-18
A	White KE, et al; Gene, 246(1-2):347-356; 2000-04-04 See the abstract.	1-18

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason(as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

15 August 2001 (15.08. 01)

Date of mailing of the international search report

30 AUG 2001 (30. 08. 01)

Name and mailing address of the ISA/

The Chinese Patent Office
6, Xitucheng Road, Haidian District,
Beijing, 100088, China

Facsimile No. 86-010-62019451

Authorized officer

ZENG, Fanhui

Telephone No. 62093733

国际检索报告

国际申请号

PCT/CN01/00749

A. 主题的分类

Int.Cl⁷: C12N15/54, C12N9/10, C07K14/435, A61K38/45

按照国际专利分类表(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类体系和分类号)

Int.Cl⁷: C12N15/54, C12N9/10, C07K14/435, A61K38/45

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称和, 如果实际可行的, 使用的检索词)

WPI, GenBank+EMBL+DDBJ+Swiss-Prot+PIR+PDB

C. 相关文件

类 型*	引用文件, 必要时, 包括相关段落的说明	相关的权利要求编号
A	EMBL AJ271385 HSA271385 2000-04-26 见序列	1-18
A	GenBank AC005833 1999-01-30 见序列	1-18
A	EMBL AL022577 HS353H6 1999-11-23 见序列	1-18
A	GenBank U91318 HUU91318 2000-01-10 见序列	1-18
A	White KE, 等; 基因, 246(1-2):347-356; 2000-04-04 见摘要	1-18

☐ 其余文件在C栏的续页中列出。☐ 见同族专利附件。

* 引用文件的专用类型:

“A” 明确表示了一般现有技术, 不认为是特别相关的文件

“E” 在先文件, 但是在国际申请日的同一日或之后公布的

“L” 对优先权要求可能产生怀疑或者用来确定另一篇引用文件的公布日期或其它特殊理由而引用的文件(如详细说明)

“O” 涉及口头公开、使用、展览或其他手段的文件

“P” 在国际申请日之前但迟于所要求的优先权日公布的文件

“T” 在国际申请日或优先权日之后公布的在后文件, 它与申请不相抵触, 但是引用它是为了理解构成发明基础的理论或原理

“X” 特别相关的文件; 当该文件被单独使用时, 要求保护的发明不能认为是新颖的或不能认为具有创造性

“Y” 特别相关的文件; 当该文件与其他一篇或多篇这类文件结合在一起, 这种结合对本领域技术人员是显而易见的, 要求保护的发明不能认为具有创造性

“&” 同族专利成员的文件

国际检索实际完成的日期

15.8 月 2001(15.08.01)

国际检索报告邮寄日期

30. 8月 2001(30.08.01)

国际检索单位名称和邮寄地址

中国专利局

中国北京市海淀区西土城路6号(100088)

传真号:

86-010-62019451

授权官员



电话号码: 010-62093733